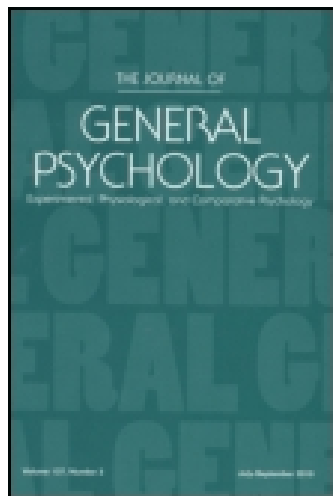




The Journal of General Psychology

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/vgen20>

Further Card-Guessing Experiments

Raymond Royce Willoughby ^a

^a Brown University , USA

Published online: 06 Jul 2010.

To cite this article: Raymond Royce Willoughby (1937) Further Card-Guessing Experiments, The Journal of General Psychology, 17:1, 3-13, DOI: [10.1080/00221309.1937.9917970](https://doi.org/10.1080/00221309.1937.9917970)

To link to this article: <http://dx.doi.org/10.1080/00221309.1937.9917970>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

FURTHER CARD-GUESSING EXPERIMENTS*

Brown University

RAYMOND ROYCE WILLOUGHBY

It has recently become possible to conduct additional experiments as a partial¹ check upon the clairvoyance hypothesis of Dr. J. B. Rhine of Duke University.²

Eight subjects were studied, of whom one (the writer) had already been studied under slightly different conditions; of these, five were studied thoroughly; it was impossible, for reasons not connected with the experiment, for the others to devote sufficient time to the work to give adequate series. The conditions approach as closely as practicable to those used by Rhine, within the limits believed necessary by the present investigator, viz.: an adequate series was defined as 200 successive trials ("runs") of 25 calls each (the guessed pack—constitution known to the subject—consisting of 5 cards of each of five suits: circle, plus, lines, rectangle, star) and all series include all the data gathered for the subject concerned. This departs from Rhine's procedure in two particulars: (*a*) no preliminary testing to establish and build up ability to score significantly was included, and (*b*) no subject was stopped because he could no longer score significantly; all scores were recorded beginning with the first trial, and all the adequately tested subjects were stopped after the 200th trial. This figure, somewhat arbitrarily determined, proved (as judged from the smoothness of the distributions and the frequency of extreme scores) to be a good effective minimum. Each

*Received in the Editorial Office on January 28, 1936.

¹Partial, because the proposition "Some subjects are psychic" cannot be tested (practically) save by first converting it to "These subjects are psychic"; since "these subjects" are either inaccessible or no longer claimed to be psychic, the best that can be done is to test some subjects who are accessible and *might* be psychic; the multiplication of such subjects would lead either to the discovery of additional psychic subjects or to the growth of doubt that the original subjects were so.

²Thanks are due to Dr. Rhine and his collaborators for the opportunity to examine at first hand the methods and some of the results of his investigations; and to Mrs. Esther D. Hunt, whose services have been employed in the gathering of the data here evaluated.

subject was given complete leeway in the mode of guessing, and several such modes appeared; e.g., a subject was free to guess slowly and laboriously, trying to "see" each card before he called it; or he could race through the 25 calls at top speed, merely saying any of the five names that occurred to him; or he could excogitate any "system" that seemed to him likely to be effective, such as naming all five of a suit in succession or grouping his calls in blocks of five and varying the order within the block. He was also given complete liberty as to the terms in which he reported, and the experimenter recorded the calls in symbols, as does Rhine. The pack was shuffled three times, edge to edge, and cut before each trial; and a "dummy" or empirical control consisting of a second shuffled but unopened pack was used for each run. All guessing was done with the pack unopened (Rhine's *DT* condition). The order of procedure for each run was: shuffle and cut dummy pack and lay it aside; shuffle and cut target pack and lay it before subject to call; record subject's calls; record target pack; record dummy pack. A sample run, the 164th by the subject *GP*, is as follows (the symbols in which the record was made being transcribed into initials of the form names: *S* refers to the subject's guesses, *T* to the target pack, and *D* to the dummy pack):

<i>S</i>	sprr	plecs	cprlp	cppss	lrler	
<i>T</i>	lppcc	srpls	spcsl	lpcrp	rlrcr	6
<i>D</i>	spscr	rscrc	rsclr	pplcp	lpsll	5

Between *S* and *T* there are six correspondences or hits, viz., on the 2nd, 10th, 12th, 17th, 24th, and 25th calls of the run, and this number is entered opposite the *T* row; between *T* and *D*, in like fashion, there are five hits. Since these hit scores are the obviously appropriate measures of the efficiency of *S* and *D* respectively in predicting *T*, it is with them that we shall be principally concerned in evaluation.

The subjects were all adult; while the investigator's scientific philosophy⁸ would not permit him to select for study only subjects who had proved their ability to score significantly, as much of an effort as was practicable was made to secure the cooperation of per-

⁸Since he suspects a possible source of error in the selection of favorable temporal sequences—rationalized in terms of the apparently reasonable but really question-begging principle mentioned.

sons who were interested or who believed they might be able to exceed chance. An endeavor was made to foster a spirit of fair trial, and as a matter of fact there was considerable interest and some rivalry on the part of the subjects, particularly during favorable sequences. A brief characterization of the subjects follows:

Subject *GP* is an unmarried man in his early twenties; his attitude was one of trying to outguess the pack, and he contrived several systems for doing so, confessing to something of a disappointment at the negative results.

Subject *RW* is the author; he is a married man of thirty-nine, skeptical in approach but desirous of giving the matter a satisfactory trial; a previous series has been reported for him, in which he acted as his own experimenter; the present series is free from any influence through seeing the previous calls of a run while making and recording the subsequent ones.

Subject *DP* is an unmarried man of forty, well known among his friends for a belief in the reality and validity of unseen forces; but his feeling after a few trials was one of slight aversion and incredulity.

Subject *MM* is a married woman in her thirties; she is the only non-academic person used, and has a reputation among her friends as a sensitive.

Subject *HP* is a married woman in her late twenties; she was very meticulous in guessing, insisting on "seeing" each card before she called it; the work was difficult for her, partly because it was so slow and largely because whenever she inquired about her success (subjects were given information as to their success—after runs only, since no one saw the target till the run was completed—when they desired it, but no standardization of the point was attempted) the results were so discrepant with the feeling of success or failure that she felt constantly thwarted.

Subject *EK* is a married woman in her middle twenties; she volunteered for the experiment and took it rather equably, finding it of no particular annoyance or interest.

Subject *KG* is an unmarried woman in her early twenties; she manifested a competitive spirit, betting small sums on her guesses as against the predictions of dummy; during unfavorable sequences she complained jocularly that since she was below chance average "the laws of chance should be repealed."

Subject *EH* was the experimenter for most of the series; she is a married woman in her late twenties, and volunteered before

the experiment that "if anyone is psychic, I am—I have hunches all the time"; the author served as experimenter for her.

Two professional psychics, one of each sex, were approached, but refused to cooperate, the man because no remuneration was available, the woman (with some indignation) because of the danger to her own powers from contamination with an alien pack of cards. Of the academic subjects, all but *DP* are professionally in contact with psychology, and all of these but *HP* have had rather extensive training in it. With the slight exceptions noted, no particular animus against the experiment was found; in fact, it is the author's impression that a fair degree of receptiveness exists among the rank and file of psychologists, while the layman is more likely impatiently to aver that he "doesn't believe in any of that stuff."

Since in counting hits there is some liability to systematic error (since a correspondence is fairly likely to be overlooked, but one is not likely to be hallucinated) all this work was checked by a second person, and most of it by a person familiar neither with what the data signified nor with the purpose of the experiment. The total number of errors discovered in the original scoring was around 10, which represented the error (in a total of approximately 7000), which would have been present due to this source without checking. For economy of space we shall present in one table (Table 1) (*a*) the subject's initials (*D* refers to the accompanying dummy); (*b*) the number of runs, *N*; (*c*) the frequency of each hit score; (*d*) the mean number of hits per run, *M*; (*e*) the probability, *P*, based upon the chi-square test for goodness of fit, that a fit as bad as or worse than the one observed could have arisen by chance. This fit refers to the correspondence between the frequency curve of hit scores for the subject and that for a theoretical chance distribution, based upon the expansion of the binominal $(p^{\frac{1}{2}} + q^{\frac{1}{2}})^{25}$; the successive ordinates of this latter curve for a population of 200 are given in the last row, and if desired the corresponding ordinates for the other populations involved (25, 161, 64) may be easily computed by merely multiplying the ordinates given by 25/200, etc.

It will be immediately clear that these results are unequivocally negative. Rhine's "*X*" value (critical ratio, essentially—deviation divided by probable error of the ideal proportion) for the largest positive deviation from chance, that of the subject *MM* (42 hits

TABLE 1

Subject	N	Hits														M	P
		0	1	2	3	4	5	6	7	8	9	10	11	12-14			
GP	200	1	5	16	31	39	40	36	20	8	3		1		4.7	.72	
D	200		8	20	22	41	44	21	21	13	6	3	1		4.8		
RW	161		4	16	20	34	33	19	16	10	6	3			4.9	.59	
D	161	2	4	10	22	33	28	20	28	7	5	1	1		4.9		
DP	25			1	3	8	3	6	2	2					5.0	.26	
D	25		1	3	2	6	4	6	3						4.6		
MM	200	1	2	15	34	31	30	33	24	15	8	3	2	2	5.2	.27	
D	200	1	6	20	31	28	39	32	18	19	2	2	1	1	4.9		
HP	200	1	7	14	27	34	41	40	14	11	7	2	1	1	4.9	.72	
D	200	2	4	16	24	34	32	34	32	9	7	5		1	5.1		
EK	200	1	6	18	27	36	41	28	15	21	3	1	2	1	4.9	.27	
D	200	1	6	12	26	40	40	37	18	14	1	2	3		4.9		
RG	200	2	6	20	28	45	27	29	18	18	4	1	1		4.8	.19	
D	200	2	6	16	31	34	30	36	24	15	5	1			4.9		
EH	64		2	2	8	11	21	13	3	2		2			4.9	.12	
D	64		2	2	12	14	9	14	7	1	2	1			4.9		
Theoretical	200	1	5	14	27	37	39	33	22	13	6	2	1	0			

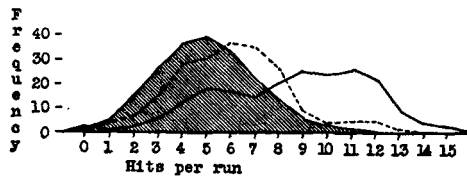
in excess of chance average) is only 2.2; and all the P values indicate that the obtained curves are fitted very well by the theoretical chance distribution.⁴

While the significance of the numerical values is reasonably clear, they will probably gain in vividness by being graphed; we select for this purpose (Figure 1a) the most nearly "psychic" subject (MM , solid line) and most nearly "psychic" dummy (HP_D , broken line) and plot the frequency distributions of their hit scores against a background of the corresponding theoretical chance distribution. With these we compare two distributions (Figure 1b) from the work of the Duke subject Hubert Pearce, derived from old records kindly placed at our disposal by Dr. Rhine and his assistants, and

⁴A failure of the conditions of the experiment to meet the assumptions of the $\sqrt{npq}$ formula, in that the successive events are not independent, was pointed out to me by Dr. Burton H. Camp.



a. Hit-score distribution of most nearly "psychic" subject (solid line) and most nearly "psychic" dummy (broken line) of present experiment, compared with theoretical chance (shaded)



b. Two hit-score distributions (lines) from work of Hubert Pearce (Duke experiments), compared with theoretical chance (shaded)

FIGURE 1

similarly plotted. It is morally certain that the distributions represented in the second figure did not arise by chance; further than to express our belief that it has not been and is not likely to be shown that they arose from occult sources, we refrain from comment upon this clear positive finding.

The temporal aspects of success or failure are important in so far as positive results may conceivably arise from selections of favorable sequences. In three cases (two of real subjects) early favorable sequences elicited considerable interest and some hope of finding a characteristically successful subject, and in some cases there were rallies at a middle or late stage of the experimentation; while unfavorable sequences were equally or perhaps more common. For each subject, therefore, the total sequence of 200 has been divided into 10 successive sub-sequences of 20 runs each, and the hit-score frequencies of these are shown (Table 2) accompanied by the corresponding figures for the dummy scores.

TABLE 2

Subject	N	Hits in successive 20's										
GP	200	107	99	96	97	88	92	89	86	113	77	
D	200	94	99	108	97	83	102	89	92	107	99	
RW	161	90	106	93	78	109	104	98	100	(8)*		
D	161	93	91	100	105	87	89	106	119	(9)		
DP	25	103	(21)*									
D	25	92	(22)									
MM	200	118	104	110	100	104	99	99	98	114	96	
D	200	90	101	105	89	84	110	86	94	104	114	
HP	200	90	109	123	96	87	93	97	87	90	112	
D	200	122	112	90	102	102	102	105	108	105	81	
EK	200	101	103	101	96	99	100	102	83	101	97	
D	200	100	95	98	98	102	103	93	113	98	88	
RG	200	95	82	90	98	73	100	106	109	117	90	
D	200	99	101	90	101	84	101	103	93	106	97	
EH	64	98	107	88	(21)*							
D	64	92	94	113	(12)							

*Incomplete 20's.

While we do not have very extreme temporal variations here, a few values are of much interest, viz., the initial 20's of *MM* and *HP_D* and the next to the last 20 of *RG*; the "X" values for these are approximately 3 (3.7 for *HP_D*) and suggest that by "practicing" a subject until a favorable run started and stopping him when it began to "fade" the total hit score might be significantly increased. It may also be observed that considerable sequences lower than chance average have occurred, e.g., with *GP* and *HP*.

No completely satisfactory solution has been found for the problem of preference, i.e., for the problem of control introduced by the fact that subjects will ordinarily not achieve an absolutely equal distribution of guesses among the five suits, but will prefer one or more at the expense of others. This results in the fact that neither the theoretical nor the empirical check, both of which are restricted to equal distribution, is an exact control for the guesses. The best technique, perhaps, would be to require the subject to match one deck with another, the faces of one or both remaining unseen. The

reasons for not doing this were several: it was desired to keep the conditions of the experiment as similar as possible to Rhine's; the matching method is much slower; and with it a rather formidable problem of controlling for secondary cues such as spots of dirt, etc., is introduced. It was thought that a sufficient number of runs would be made available so that a sub-population consisting of only no-preference runs could be drawn and studied; but no such sub-population appeared in sufficient size to yield stable conclusions. The appropriate preference measure is the sum of deviations from five, e.g., in a run of guesses containing six circles, four lines, five plus, seven rectangles, and three stars the preference measure would be six. The distribution of such measures is as shown in Table 3.

TABLE 3

Subject	N	preference index						
		0	2	4	6	8	10	12
GP	200	30*	76	74	18	2		
RW	161	8	43	80	24	5	1	
DP	25		4	13	3	4	1	
MM	200	3	37	74	59	19	7	1
HP	200	8	55	93	38	5	1	
EK	200	3	39	77	58	19	3	1
RG	200	2	49	83	45	18	1	2
EH	64	2	23	26	8	5		
All	1250	56	326	520	253	77	14	4
%	100	4.5	26	42	20.2	6	1	.7

*This relatively high figure is due to the use of "systems" by this subject

It is thus quite unusual for any subject (unless "working a system") to achieve an equal distribution of suits; the modal distribution of guesses among the five suits is of the type 4 5 6 6 4 (index 4), and nearly nine-tenths of the runs have a preference index of 2, 4, or 6; there are also large individual differences, particularly with respect to the more extreme inequalities of distribution (e.g., *MM* and *HP*). For each subject, the preference index was plotted against the hit score (for each run), and in every case there was clear absence of relationship; departures from equal distribution are therefore without effect on success in guessing.

Other aspects of preference are without special bearing on the problem of *ESP*, but are of some interest in themselves. The total figure preference, or distribution among the figures of all the guesses

made by a given subject, was, unexpectedly, fairly even; in the case of no subject did the preference for any figure depart from 20 per cent by more than 3 per cent, and for all subjects together the figures for the circle, lines, plus, rectangle, and star were (in percentages) 21.4, 20.0, 20.3, 19.1, and 19.2 respectively. We may perhaps attribute this result to the constant self-administered *Aufgabe* to maintain an equal distribution, which would gain in effectiveness over a large number of calls (5000 in the case of the major subjects).

The "start preferences," or preferences shown in the first call of each run of 25, show marked differences both among subjects and among figures; they are given below (Table 4) in the usual

TABLE 4

Subject	N	c	l	p	r	s	TD
GP	200	44	5	12	18	21	50
RW	161	18	28	19	23	12	22
DP	25	24	20	12	20	24	16
MM	200	41	5	3	23	28	64
HP	200	47	33	5	8	7	80
EK	200	22	12	18	26	22	20
RG	200	25	11	33	16	15	36
EH	64	30	12	11	8	39	58
All	1250	33	15	15	18	19	26

order (in percentages), with the addition of a column (*TD*) showing the total deviation of each subject from the equal-distribution figure of 20 per cent.

The circle is thus preferred for the initial call about twice as often as any of the other figures, and the individual differences range from a fairly even distribution, in which the most preferred figure is about twice as frequent as the least preferred (*EK*), to one in which it is over nine times as frequent (*HP*).

The problem of preferential sequence is one which might be given extended analysis, and might even prove to have some bearing upon the major problem, inasmuch as (if the cards were badly shuffled) characteristic individual guessing sequences might coincide with persisting sequences in the cards and so spuriously raise the score.⁵ It

⁵Thus in the unsystematized "playing around" with the cards which was the usual immediate reaction to Dr. Rhine's first monograph, the writer encountered a young lady unaccustomed to handling cards; she made a good but not extreme score on the first run, and was then permitted to shuffle

was thought sufficient in the present connection, however, merely to demonstrate the presence of preferential sequence, and for this purpose the short series given by the subject *DP* was completely analyzed to two positions only; i.e., the figure following each figure was considered, but the figure following that was not. The results are given in Table 5, which is to be read "*c* (circle) was followed by *c*

TABLE 5

		c	l	p	r	s	0	T
c		24	32	18	42	20	7	143
	%	18	24	13	31	14		
l		28	21	37	22	25	8	141
	%	22	16	28	16	19		
p		28	24	21	22	15	6	116
	%	25	22	19	20	14		
r		38	31	10	14	19	2	114
	%	34	28	9	12	17		
s		19	28	26	10	32	2	107
	%	19	27	24	9	21		

24 times, by *l* (lines) 32 times, . . . and by nothing (i.e., was the end of a run) 7 times." In computing the percentages the base used did not, of course, include the 0's.

The reversible configuration *cr-rc* may be mentioned in passing; it appears that these common geometrical figures are associated with each other (at least in the mind of this subject) substantially more closely than any of the other figures is associated with them. For the rest, it seems clear that preferential sequence exists in the series; in the extreme case (*r*) a figure may be associated about four times as often with one figure as with another.

We have still to face the problem raised by the discrepancy between these results and Rhine's. We believe that the latter are due to an undetected artifact, the nature of which we cannot demonstrate since we have not observed the gathering of data by the Duke methods from "psychic" subjects. The strictly correct control for

the cards herself, which she did by holding them on edge with the left hand and with the right raising a few from the middle region and replacing them, supposedly in changed positions but really very much in the same order. She then guessed again (maintaining more or less characteristic sequences) and made a score of 16, which would occur by pure chance about once in 5,000,000 runs; the "ability," of course, promptly disappeared when the cards were thoroughly shuffled.

the factor of preference has not been suggested, but if it be admitted that control by the binomial expansion is approximately satisfactory, then the further necessary controls are two, viz., adequate shuffling of the cards (minimum of three edge-to-edge shuffles and a cut, by the experimenter), and sufficiently long unselected series (minimum of 200 consecutive unselected trials under the same conditions for each subject, without preliminary test series or curtailment for "loss" of the ability). It is possible that experience will prove these standards are more strict than is necessary; but it is certain that any series surviving them will be entitled to further consideration, and not at all certain that any series not surviving them is so entitled. We do not believe that they have been met save in our own experiments; if and when they are, we are convinced that the results, like ours, will be explicable by the hypothesis of chance alone.

Brown University
Providence, Rhode Island